

Einführung in IR - Winter 2006/07

Dipl.-Inform. Sascha Kriewel, LF 137

sascha.kriewel@uni-due.de

Übungsblatt 2

**Evaluierung
keine Abgabe**

Aufgabe 3: Mittelwertberechnung für Recall und Precision

In der Vorlesung hast Du gehört, dass es zwei verschiedene Möglichkeiten zur Mittelwertbildung gibt:

Makrobewertung

$$p_M = \frac{1}{N} \sum_{i=1}^N \frac{|REL_i \cap GEF_i|}{|GEF_i|}$$

und Mikrobewertung.

$$p_\mu = \frac{\sum_{i=1}^N |REL_i \cap GEF_i|}{\sum_{i=1}^N |GEF_i|}$$

Zu welchem Zweck ist die Mittelwertbestimmung bei Recall und Precision nötig oder sinnvoll? Warum benötigt man die zwei Ansätze? Stelle die analogen Formeln für den Recall auf.

Aufgabe 4: Bestimmung von Precision und Recall

Nimm an, Du suchst für ein Kurzreferat Quellen über die Intelligenz von verschiedenen Meeressäugtieren, insbesondere im Vergleich zu Primaten. Formuliere hierzu eine Anfrage und benutze die gleiche Anfrage in zwei Websuchmaschinen (z.B. Google.com, Yahoo.com, MSN.com, ...).

Betrachte jeweils die ersten 10 Dokumente des Ergebnisses und beurteile sie bezüglich ihrer Relevanz zur Suchanfrage. Gehe im Folgenden nun davon aus, dass die Vereinigung der relevanten Dokumente aus den beiden betrachteten Ergebnismengen die Gesamtmenge der relevanten Dokumente ist. (Was müsste man tun, um zu einer besseren Abschätzung des Recalls zu kommen?)

Bestimme Recall und Precision an den einzelnen Rängen der beiden Ergebnisse und trage die Wertepaare als Punkte in einen gemeinsamen Graphen ein. Wie kann man nun die Qualität der beiden Ergebnisse vergleichen?

Aufgabe 5: Precision vs. Recall

Precision und Recall sind zwei wichtige Retrievalmaße. Nicht immer ist es möglich, Precision und Recall gleichermaßen zu verbessern. Überlege Dir jeweils Einsatzbeispiele für Text- oder Bildretrieval, in denen Du eher eine hohe Precision oder eher einen hohen Recall bevorzugen würdest.